ЗАДАЧА ЛОКАЛЬНОГО АВТОМАТИЧЕСКОГО ОБУЧЕНИЯ В АЛГЕБРАИЧЕСКИХ БАЙЕСОВСКИХ СЕТЯХ: ЛОГИКО-ВЕРОЯТНОСТНЫЙ ПОДХОД

А. Л. Тулупьев

Санкт-Петербургский институт информатики и автоматизации РАН

СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

<alt@iias.spb.su>

УДК 004.8

Тулупьев А. Л. **Задача локального автоматического обучения в алгебраических байесовских сетях: логико-вероятностный подход** // Труды СПИИРАН. Вып. 7. — СПб.: Наука, 2008.

Аннотация. *Одной из проблем, сдерживающих развитие интеллектуальных информационных систем, является дефицит знаний (knowledge bottleneck). Среди перспективных способов его преодоления — решение задачи машинного обучения для моделей представления знаний с неопределенностью, которые используются в интеллектуальных системах. Алгебраические байесовские сети — одна из таких вероятностных графических моделей, отличающаяся от остальных тем, что позволяет представлять и обрабатывать интервальные оценки вероятности истинности. Цель работы — описать задачу автоматического обучения в отношении фрагмента знаний алгебраической байесовской сети, а также предложить пути ее решения и указать трудности, с которыми приходится сталкиваться.* — Библ. 16 назв.

UDC 004.8

Tulupyev A. L. **A Local Machine Learning Task in Algebraical Bayesian Networks: a Probabilistic-Logic Approach** // SPIIRAS Proceedings. Issue 7. — SPb.: Nauka, 2008.

Abstract. *One of the problems that slow down intelligent information systems development and industry-wide spread is so-called knowledge bottleneck. Machine learning for various uncertain knowledge representations and models used in intelligent systems is a promising way to cope with the bottleneck. Algebraical Bayesian networks are a probabilistic graphical model that allows for representing and processing interval estimates of probabilities. The paper goal is to describe the machine learning task in regard to a knowledge pattern of an algebraical Bayesian network as well as to present a few ways to solve the task and to outline obstacles related to those ways.* — Bibl. 16 Items.

1. Введение. Дефицит знаний

Хотя проблема промышленного внедрения информационных систем, использующих и способных обрабатывать данные и знания с неопределенностью, становится все более актуальной, ее решению мешают два так называемых «узких места»: 1) дефицит моделей для представления данных и знаний с неопределенностью (uncertain knowledge representation bottleneck), 2) дефицит данных и знаний (knowledge bottleneck) [14].

Основание для дефицита первого вида достаточно очевидно: как в формальных системах представить и систематически обрабатывать неопределенность (недоопределенность) наших знаний, с которой зачастую приходится сталкиваться на практике. Тривиальные решения, которые сводятся просто к тому, чтобы отбросить знания (или данные) с неполнотой, непригодны по двум причинам: 1) отнюдь не во всех предметных областях за разумное время можно получить «полные знания»; 2) человек способен действовать и в условиях неполноты знаний о складывающейся ситуации (не без того, конечно, чтобы перед принятием решений не попытаться уменьшить степень своей неосведомленности).

Дефицит второго вида заметно менее очевиден, но еще более остр. Есть предметные области, где просто не существует человека-эксперта (можно быть уверенным, что еще ни один живой специалист не смог поработать в жерле действующего вулкана или на поверхности Венеры, но роботизированные комплексы в такие места посылать требуется). Кроме того, не все эксперты готовы делиться своим знанием. Наконец, при попытках организовать автоматическое обучение, машинное обучение (machine learning, на самом деле — автоматический подбор некоторых моделей связей между случайными элементами, характеризующими предметную область, а также оценку параметров этих моделей) по данным из баз данных вдруг обнаруживается, что эти данные неполны (то есть имеют пропуски) и частью или зачастую недостоверны (например, неисправное средство измерения (СИ) выдает неправильный результат, перепутаны измерительные каналы, СИ установлено на позиции, которая не отвечает ряду требований, и прочее).

Цель работы — описать задачу автоматического обучения в отношении фрагмента знаний алгебраической байесовской сети [3–10], а также предложить пути ее решения и указать трудности, с которыми приходится сталкиваться. Кроме того, будут сформулированы основные направления дальнейших исследований (включая некоторые виды систематизации), связанные с вопросами автоматического обучения алгебраических байесовских сетей.

2. Содержание задачи автоматического обучения в случае вероятностных графических моделей

В первую очередь предполагается, что знания о предметной области могут быть представлены в виде системы случайных элементов.

В качестве исходных данных рассматривается выборка совместных означиваний сразу всех наблюдающихся случайных элементов. Эту выборку иногда называют *обучающей выборкой*. (В некоторых разделах искусственного интеллекта мы еще сталкиваемся с *тестирующей выборкой*.)

Кроме того, эксперты могут предоставить дополнительные сведения о связях случайных элементов и о распределениях вероятностей над некоторой частью случайных элементов или над всеми ними.

В понятие автоматического обучения включается процессы 1) построения структуры системы (модели, графической модели) и 2) подбора значений параметров для такой системы (модели). Предполагается, что компьютерная программа автоматически, уже без вмешательства специалистов, достигает указанных результатов, обрабатывая доступные исходные данные, сведения, знания.

Автоматическое обучение может быть также названо машинным обучением, автоматическим машинным обучением, и просто обучением. Для краткости в работе мы будем придерживаться последнего термина.

Заметим, что постановка задачи автоматического обучения дана в самом общем виде. Процессы, объекты и цели обучения, а также исходные данные в каждом конкретном классе вероятностных графических моделей могут быть гораздо более глубоко диверсифицированы.

3. Локальные задачи автоматического обучения в случае алгебраических байесовских сетей

Процесс построения структуры (более точно — вторичной структуры) алгебраической байесовской сети по имеющимся сведениям о фрагментах знаний описан в [9]. Он нацелен на решение одной из задач глобального обучения АБС. Целью же настоящей статьи является рассмотрение задач, возникающих при локальном обучении АБС.

В случае локального обучения речь идет о формировании отдельного фрагмента знаний — идеала конъюнктов со скалярными или интервальными оценками вероятности их истинности. (В рассмотрение также удобно включить фрагмент знаний, образованный над набором квантов [7].) Структура фрагмента знаний определяется числом атомарных пропозициональных формул (атомов) которые в него попали. Таким образом, локальное обучение сводится преимущественно к одной задаче — к формированию набора оценок вероятности истинности над фрагментом знаний с известной структурой по исходному набору совместных означиваний атомарных литералов. Для фрагмента знаний, построенного над идеалом конъюнктов, набор оценок вероятностей записывается в виде вектора $\mathbf{P}$; для фрагмента знаний над набором квантов — в виде вектора $\mathbf{Q}$[1]. Эти два вектора оценок тесно связаны между собой [5, 7, 8, 10, 11].

Как уже отмечалось, в теории автоматического обучения исходный набор означиваний принято называть обучающей выборкой (или просто выборкой). Этот термин вполне закономерно унаследован из математической статистики. Элементы выборки с точки зрения теории АБС являются квантами.

Пусть задан фрагмент знаний над алфавитом из трех атомов

$$A = \{x_2, x_1, x_0\}.$$

Набор квантов

$$S = \{x_2 x_1 x_0, x_2 \bar{x}_1 x_0, x_2 x_1 x_0, x_2 \bar{x}_1 \bar{x}_0, \bar{x}_2 x_1 x_0, x_2 x_1 x_0, x_2 \bar{x}_1 x_0\}$$

служит примером выборки.

Для каждого отдельно взятого фрагмента знаний алфавит фиксирован, а порядок следования атомов в кванте может быть заранее оговорен, поэтому нет необходимости все время использовать буквенно-индексные обозначения для того, чтобы указать означивания того или иного атомарного литерала. Договоримся его отрицательному означиванию сопоставлять 0, а положительному — 1. Тогда та же самая выборка может быть записана как

$$S = \{111, 101, 111, 100, 011, 111, 101\},$$

что заметно короче. Еще несколько примеров обучающей выборки приведено в табл. 1.

4. Фрагмент знаний с бинарными оценками вероятностей

Заметим, что предложенные значения 0 и 1 можно трактовать как вероятность положительного означивания случайного бинарного элемента, соответствующего атому. При таких вероятностях положительного означивания (а именно при вероятностях со значениями 0 или 1) однозначно определяется вероятность любого конъюнкта, то есть конъюнкции некоторого набора атомов, а

[1] Здесь и далее для краткости используются обозначения: $\mathbf{Q} = \mathbf{P_q}$ и $\mathbf{P} = \mathbf{P_c}$ [8, 11].

Труды СПИИРАН. 2008. Вып. 7. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2008. Issue 7. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

именно, вероятность конъюнкта $x_{n-1}\dots x_1x_0$ можно получить по любой из двух формул:

$$p\left(x_{n-1}\dots x_1x_0\right)=p\left(x_{n-1}\right)\dots p\left(x_1\right)p\left(x_0\right), \tag{Y1}$$

$$p\left(x_{n-1}\dots x_1x_0\right)=\min\left\{p\left(x_{n-1}\right),\dots p\left(x_1\right),p\left(x_0\right)\right\}. \tag{Y2}$$

Более того, справа в формулах (Y1–Y2) может использоваться любая триангулярная норма, вместо таковых, заданных максимумом и произведением. Однако формулами (Y1–Y2) уже нельзя будет воспользоваться, если вероятность двух или более атомов не будет равна 0 или 1. В таком случае даже нельзя ожидать, что оценка вероятности останется скалярной; скорее всего, придется пользоваться интервальными оценками, если не прибегать к дополнительным предположениям. Это — существенный момент.

Таблица 1

Вероятности конъюнктов в ФЗ над одним атомом

Означивание квантов	Вероятность конъюнктов
	$p(x_0)$
$\bar{x}_0$	0
x_0	1

Таблица 2

Вероятности конъюнктов в ФЗ над двумя атомами

Означивание квантов	Вероятность конъюнктов		
	$p(x_0)$	$p(x_1)$	$p(x_1x_0)$
$\bar{x}_1\bar{x}_0$	0	0	0
$\bar{x}_1x_0$	1	0	0
$x_1\bar{x}_0$	0	1	0
x_1x_0	1	1	1

Таблица 3

Вероятности конъюнктов в ФЗ над тремя атомами

Означивание квантов	Вероятность конъюнктов						
	$p(x_0)$	$p(x_1)$	$p(x_1x_0)$	$p(x_2)$	$p(x_2x_0)$	$p(x_2x_1)$	$p(x_2x_1x_0)$
$\bar{x}_2\bar{x}_1\bar{x}_0$	0	0	0	0	0	0	0
$\bar{x}_2\bar{x}_1x_0$	1	0	0	0	0	0	0
$\bar{x}_2x_1\bar{x}_0$	0	1	0	0	0	0	0
$\bar{x}_2x_1x_0$	1	1	1	0	0	0	0
$x_2\bar{x}_1\bar{x}_0$	0	0	0	1	0	0	0
$x_2\bar{x}_1x_0$	1	0	0	1	1	0	0
$x_2x_1\bar{x}_0$	0	1	0	1	0	1	0
$x_2x_1x_0$	1	1	1	1	1	1	1

Таким образом, каждому кванту можно сопоставить фрагмент знаний с бинарными оценками вероятности истинности, то есть с оценками вероятности,

Труды СПИИРАН. 2008. Вып. 7. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2008. Issue 7. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

которые могут принять одно из двух значений: 0 или 1. В табл. 1–3 фрагменты знаний небольшого порядка сопоставлены квантам.

Вероятность пустого конъюнкта нигде не указана; она всегда равна 1.

5. Обучение по выборке: означивания без пропущенных элементов

Последовательность означиваний, из которых состоит выборка, нелегка для восприятия и труднообозрима, особенно если она большая. Однако выборку можно записать в более компактной форме: кванты упорядочить, повторы среди них исключить, каждому кванту сопоставить число — сколько раз он встречается в обучающей выборке. Если поделить каждое такое число на общее число элементов в выборке, то получим оценку вероятности истинности кванта. Таким образом будет сформирован вектор $\mathbf{Q}$, задающий распределение вероятностей на квантах. Его можно преобразовать в вектор $\mathbf{P}$ — особым образом упорядоченную совокупность вероятностей конъюнктов — с помощью уравнения $\mathbf{P} = \mathbf{JQ}$ [5, 7, 8, 10, 11].

Любой из векторов $\mathbf{Q}$ и $\mathbf{P}$ может служить результатом локального обучения по имеющейся выборке. Как правило, число элементов во фрагменте знаний невелико; много меньше, чем объем выборки; поэтому получившийся результат будет более обозрим, чем исходные данные.

Таблица 4

Примеры обучающих выборок S_i без пропусков

№	Означивание одного литерала		Означивание двух литералов			Означивание трех литералов		
	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
0	0	1	01	11	11	000	011	101
1	0	1	00	01	01	010	001	110
2	1	1	01	01	11	101	001	101
3	0	0	00	11	11	111	110	110
4	1	0	00	10	11	111	111	101
5	1	1	00	01	01	010	111	100
6	0	1	10	01	01	011	010	100
7	0	1	10	01	01	110	010	110
8	0	1	11	11	11	101	000	001
9	1	1	11	00	00	111	000	010

Формализуем описанную процедуру и обсудим ее некоторые особенности, незаметные на первый взгляд.

Пусть выборка состоит из набора квантов $S = \left\{q^i\right\}_{i=0}^{N-1}$, где N — это размер выборки. Каждому кванту можно сопоставить $\mathbf{Q}^i$ — вектор вероятностей квантов, построенных над соответствующим алфавитом. В векторе $\mathbf{Q}^i$ все элементы нули, за исключением того, который соответствует кванту q^i: этот элемент вектора равен единице.

При введенных обозначениях и соглашениях процедура обучения выглядит совсем просто:

$$\mathbf{Q} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{Q}^{i}\ . \qquad \text{(Y3)}$$

Такие вычисления очень удобно реализовать с использованием массивов небольшой размерности. Кроме того, формулу (Y3) легко адаптировать для ситуации, когда обучение происходит постепенно, то есть элементы выборки доступны не все сразу, а поступают друг за другом. При этом требуется на каждом шаге иметь «обученный» уже поступившими сведениями вектор $\mathbf{Q}$. Пусть у нас совершается i-й шаг: $\mathbf{Q}[i-1]$ — результат обучения вектора $\mathbf{Q}$ предшествующими элементами выборки, и поступает элемент $\mathbf{Q}^{i}$. Тогда рекуррентная процедура «дообучения» будет выглядеть как

$$\mathbf{Q}[i] = \frac{i\mathbf{Q}[i-1]+\mathbf{Q}^{i}}{i+1} = \mathbf{Q}[i-1] + \frac{\mathbf{Q}^{i}-\mathbf{Q}[i-1]}{i+1}. \qquad \text{(Y4)}$$

Для простоты примем $\mathbf{Q}[-1]=\mathbf{0}$, поскольку на первой итерации он умножается на ноль и далее роли не играет.

Аналогичные соотношения можно записать и для вектора вероятностей конъюнктов:

$$\mathbf{P} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{P}^{i}, \qquad \text{(Y5)}$$

$$\mathbf{P}[i] = \frac{i\mathbf{P}[i-1]+\mathbf{P}^{i}}{i+1} = \mathbf{P}[i-1] + \frac{\mathbf{P}^{i}-\mathbf{P}[i-1]}{i+1}. \qquad \text{(Y6)}$$

Для удобства будем считать, что $\mathbf{P}[-1]=\dot{\mathbf{0}}$, то есть все компоненты этого вектора нулевые, за исключением самой первой: она равна 1. На данном этапе с формальной точки зрения выбор исходного вектора $\mathbf{P}[-1]$ несуществен: он все равно умножается на ноль. Однако в дальнейшем нам потребуется, чтобы этот вектор был бы непротиворечив.

Уравнения (Y4) и (Y6) можно переписать несколько иначе: так, чтобы четко было видно, что они определяют линейную комбинацию участвующих в них векторов.

$$\mathbf{Q}[i] = \frac{i\mathbf{Q}[i-1]+\mathbf{Q}^{i}}{i+1} = \frac{i}{i+1}\mathbf{Q}[i-1] + \frac{1}{i+1}\mathbf{Q}^{i}, \qquad \text{(Y7)}$$

$$\mathbf{P}[i] = \frac{i\,\mathbf{P}[i-1]+\mathbf{P}^{i}}{i+1} = \frac{i}{i+1}\mathbf{P}[i-1] + \frac{1}{i+1}\mathbf{P}^{i}. \qquad \text{(Y8)}$$

Заметим, что векторы $\mathbf{Q}[i-1]$, $\mathbf{P}[i-1]$, $\mathbf{Q}^{i}$, $\mathbf{P}^{i}$ непротиворечивы. В теории АБС известно, что линейная комбинация непротиворечивых фрагментов знаний одинаковой структуры будет являться непротиворечивым фрагментом знаний; это же справедливо и для соответствующих им векторов оценок. Таким образом, в процессе обучения у нас формируются непротиворечивые фрагменты знаний, то есть оказываются непротиворечивыми наборы оценок $\mathbf{Q}[i]$, $\mathbf{P}[i]$.

Уравнения, записанные в виде (Y7) и (Y8), можно рассматривать как один из вариантов *адаптации* нашей системы знаний $\mathbf{Q}[i-1]$ и $\mathbf{P}[i-1]$ к поступившим новым сведениям $\mathbf{Q}^{i}$ и $\mathbf{P}^{i}$. Кроме того, адаптивное обучение может строится и на иных (но похожих) уравнениях; например, на таких:

$$\mathbf{Q}[i] = \alpha\mathbf{Q}[i-1] + (1-\alpha)\mathbf{Q}^{i}, \qquad \text{(Y9)}$$

Труды СПИИРАН. 2008. Вып. 7. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2008. Issue 7. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

$$\mathbf{P}[i] = \alpha \mathbf{P}[i-1] + (1-\alpha)\mathbf{P}^i, \qquad (Y10)$$

где $0 \le \alpha \le 1$, а векторы $\mathbf{Q}[-1]$ и $\mathbf{P}[-1]$ непротиворечивы и представляют наши априорные знания или предположения о системе. Подход, представленный уравнениями (Y9) и (Y10), имеет очень глубокие теоретические основания, с ними можно ознакомиться, например, в [14].

Заметим, что выбор коэффициента α сказывается на скорости «забывания» априорного знания и на скорости «адаптации» к новым данным. Уравнения адаптации могут быть составлены и иначе; в частности, в них может комбинироваться вектор, представляющий априорное знаний, и вектор $\mathbf{Q}[N-1]$ (или $\mathbf{P}[N-1]$ соответственно). В этом случае адаптация «происходит сразу» по отношению ко всей обучающей выборке; каждый элемент выборки внесет одинаковый вклад в получившийся результат:

$$\mathbf{Q}_{\text{result}} = \alpha \mathbf{Q}_{\text{a priori}} + (1-\alpha)\mathbf{Q}[N-1], \qquad (Y11)$$

$$\mathbf{P}_{\text{result}} = \alpha \mathbf{P}_{\text{a priori}} + (1-\alpha)\mathbf{P}[N-1]. \qquad (Y12)$$

Выбор стратегии обучения (уравнений обучения) зависит от предметной области и от стоящих перед интеллектуальной системой задач. В качестве основного объекта анализа мы будем рассматривать (Y3) и (Y5), поскольку они наиболее просты, полностью отражают «идею» автоматического обучения, а также имеют обобщение на более сложные случаи, не затуманенное обилием важных, но все-таки технических деталей.

6. Означивания с пропущенными элементами

Обучение по «хорошей» выборке, в которой нет элементов с пропусками в означиваниях, идеологически просто — оно сводится к подсчету частот появления того или иного означивания. Затем полученные результаты могут быть дополнительно скомбинированы с некоторым объектом, представляющим априорные знания, однако и здесь проблем не возникает. К сожалению, не всегда оказывается так, что обучение проходит по «хорошим» (то есть без пропусков) исходным данным. Общим случаем обработки данных «с дефектами» и восстановления их пропущенных элементов занимаются целые отрасли статистики и искусственного интеллекта [1, 12, 14]. Мы же рассмотрим возникающие проблемы в рамках и на языке логико-вероятностного подхода, которого придерживается теория алгебраических байесовских сетей.

В табл. 5 приведены примеры обучающих выборок с пропусками. В них знак * стоит на месте пропущенного элемента данных (заметим, что можно выбрать и другое обозначение, например U от uncertain, undefined, undetected, либо NULL, либо -, либо ^, либо еще какой-нибудь символ).

Предложенные в табл. 5 выборки демонстрируют разную степень «некачественности» доступных сведений для обучения фрагментов знаний.

Для обучения ничего не дадут примеры S_{11} и S_{13}. Кроме числа реализаций, они не позволят определить вообще ничего.

Выборка S_{14} более информативна, однако в ней не встречается ни одного элемента, который бы содержал означивания обоих атомарных литералов. По этой выборке мы можем строить предположения о вероятности истинности каждого отдельного атома, но не о вероятностной связи их совместных означиваний. Аналогичная ситуация наблюдается и в случае выборки S_{16}.

Труды СПИИРАН. 2008. Вып. 7. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2008. Issue 7. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Таблица 5

Примеры обучающих выборок с пропусками

№	Означивание одного литерала		Означивания двух литералов			Означивания трех литералов		
	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}
0	*	1	**	*1	11	*0*	*11	1*1
1	*	1	**	*1	01	*1*	0*1	110
2	*	*	**	*1	*1	**1	00*	101
3	*	0	**	1*	11	1**	*10	*10
4	*	0	**	*0	1*	111	1*1	101
5	*	*	**	0*	0*	**0	*11	10*
6	*	1	**	*1	01	0**	0*0	100
7	*	*	**	*1	01	**0	0*0	1*0
8	*	1	**	1*	11	*0*	00*	001
9	*	1	**	0*	00	**1	0*0	010

В выборке S_{17} более тонкий дефект. На ее основе можно строить оценки вероятности истинности атомов, их попарных конъюнкций, однако в ней не встречается ни одной реализации, по которой мы могли бы судить о связи означиваний сразу всех трех атомарных литералов.

Выборки S_{15} и S_{18} отличаются от остальных в лучшую сторону: в них меньше пропусков и присутствуют элементы с известными означиваниями сразу всех атомарных литералов.

Наличие или отсутствие закономерности, которой подчиняются места расположения пропущенных означиваний, играет существенную роль. Классификация такого рода закономерностей предлагается в соответствующей литературе [1, 13, 15].

Таблица 6

Выборки без пропусков, соответствующие (не противоречащие) выборке с пропусками

№	Выборки								
	исходная	возможные							
	S_9	S_{90}	S_{91}	S_{92}	S_{93}	S_{94}	S_{95}	S_{96}	S_{97}
0	11	11	11	11	11	11	11	11	11
1	01	01	01	01	01	01	01	01	01
2	*1	01	01	01	01	11	11	11	11
3	11	11	11	11	11	11	11	11	11
4	1*	10	10	11	11	10	10	11	11
5	0*	00	01	00	01	00	01	00	01
6	01	01	01	01	01	01	01	01	01
7	01	01	01	01	01	01	01	01	01
8	11	11	11	11	11	11	11	11	11
9	00	00	00	00	00	00	00	00	00
Пропущенные означивания		000	001	010	011	100	101	110	111

Чтобы двинуться дальше, требуется исследовать семантику * — пропущенного означивания. На самом деле, на месте знака * либо мог стоять 0, либо могла стоять 1. Таким образом, выборке с пропущенными означиваниями может соответствовать несколько различающихся *выборок без пропущенных значений* (или *возможных выборок*, или *выборок, не противоречащих исходной*, или *выборок, соответствующих исходной*). В табл. 6 приведен соответствующий пример.

Если сохранять порядок следования элементов выборки, то в нашем случае при k пропущенных означиваниях существует 2^k выборок без пропусков, соответствующих (не противоречащих) исходной. Можно говорить о семействе выборок без пропусков, не противоречащих исходной выборке с пропусками. Таким образом, выборка с пропусками задает семейство распределений вероятностей, а значит, фрагмент знаний, построенный на ее основе, не может содержать лишь скалярные оценки истинности. Без дополнительных предположений, гипотез или соглашений невозможно оправдать выбор одного распределения вероятностей из многих, для того чтобы обеспечить каждый элемент фрагмента знаний скалярной оценкой вероятности истинности. Следовательно, открывается два выбора: 1) либо допустить нескалярные оценки истинности, 2) либо предложить принцип выбора или формирования одного распределения при условиях, что имеются несколько возможных.

7. Обучение интервальных оценок в условиях неопределенности

Знаку * можно дать две разные интерпретации. Ему можно сопоставить либо множество их двух элементов $\{0;1\}$, либо замкнутый промежуток $[0;1]$. В первом случае подразумевается, что на позиции, занимаемой знаком *, должен был бы стоять либо 0, либо 1. Во втором случае предполагается, что запись со знаком * лучше всего представляется (замещается) набором записей, часть которого содержит 0 вместо *, а другая часть — 1 вместо *. А анализ соотношения этих частей позволяет говорить о вероятности положительного (то есть 1) означивания той позиции, где в исходном данном стоит *; причем эта вероятность как раз может принять любое значение из $[0;1]$.

Как уже было отмечено, мы не можем ожидать, что по выборке с пропусками нам удастся получить фрагмент знаний со скалярными оценками истинности, не прибегая к дополнительным предположениям. Однако мы можем попытаться построить (обучить) интервальные оценки, если всякий раз, когда встречается знак *, будем рассматривать оба допустимых случая его означивания (0 и 1) и вести подсчет нижней и верхней оценки частоты встречаемости положительного означивания. В табл. 7 рассмотрен простейший случай — обучение оценки вероятности истинности одного атома x_0.

Точно такой же подход можно применить и при обучении вероятностей конъюнктов во фрагменте знаний, образованном над алфавитом из двух или более атомов, когда поступающая выборка содержит пропущенные элементы в означиваниях. Интерпретации поступающих означиваний для фрагмента знаний над алфавитом из двух элементов $\{x_1, x_0\}$ приведены в табл. 8. Следует заметить, что здесь мы вторгаемся в богатую область исследований, связанную со многозначными логиками [2].

Используя табл. 8 и принципы, по которым составлена табл. 9, мы можем обработать обучающую выборку и получить верхнюю и нижнюю оценку вероятности каждого конъюнкта из фрагмента знаний. Равно мы можем получить верхние и нижние оценки вероятностей квантов. Следует учесть, что если в векторах $\mathbf{Q}$ и $\mathbf{P}$ присутствуют интервальные оценки, то уравнения $\mathbf{Q} = \mathbf{IP}$ и $\mathbf{P} = \mathbf{JQ}$ не будут уже иметь места в общем случае. Потребуется решать задачи

Труды СПИИРАН. 2008. Вып. 7. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2008. Issue 7. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

линейного программирования; причем при переходах от одного вектора к другому часть информации может утрачиваться.

Таблица 7

Обучение одного параметра по выборке с пропусками

№	Наблюдаемое значение	Интерпретация	Подсчет положительных означиваний
0	1	[1;1]	[1;1]
1	1	[1;1]	[2;2]
2	0	[0;0]	[2;2]
3	*	[0;1]	[2;3]
4	1	[1;1]	[3;4]
5	0	[0;0]	[3;4]
6	*	[0;1]	[3;5]
7	*	[0;1]	[3;6]
8	0	[0;0]	[3;6]
9	1	[1;1]	[4;7]
Оценка $p(x_0)$ по частоте положительного означивания			[0.4; 0.7]

Таблица 8

Связь означиваний квантов в выборке и конъюнктов

Означивание	Интерпретация		
x_1, x_0	x_0	x_1	$x_1 x_0$
**	[0;1]	[0;1]	[0;1]
*0	[0;0]	[0;1]	[0;0]
*1	[1;1]	[0;1]	[0;1]
0*	[0;1]	[1;1]	[0;0]
00	[0;0]	[0;0]	[0;0]
01	[1;1]	[0;0]	[0;0]
1*	[0;1]	[1;1]	[0;1]
10	[0;0]	[0;1]	[0;0]
11	[1;1]	[1;1]	[1;1]

Таблица 9

Поэлементное интервальное сложение по Минковскому (примеры)

	[0;0]	[0;1]	[1;1]
[0;0]	[0;0]	[0;1]	[1;1]
[0;1]	[0;1]	[0;2]	[1;2]
[1;1]	[1;1]	[1;2]	[2;2]

Умножение и деление по Минковскому на скаляр производится поэлементно: $\alpha[a;b] = [\alpha a;\alpha b]$.

Если нас устраивают интервальные оценки, то задачи автоматического обучения по выборке с пропущенными элементами означиваний в случае алфавита из трех или более атомов решаются аналогичным образом. Однако интервальные оценки могут нас и не устраивать.

8. Обучение точечных оценок в условиях неопределенности

В этом разделе подход, предложенный Н. В. Хованым [11], адаптируется для применения к объектам теории алгебраических байесовских сетей.

Чтобы получить скалярные оценки элементов фрагмента знаний, построенного либо над идеалом конъюнктов, либо над набором квантов, необходимо выбрать какое-то одно распределение вероятностей, причем в условиях дефицита информации, то есть в случае, когда обучающая выборка содержит означивания с пропусками элементов. Такой выбор может быть осуществлен на основе явных или неявных дополнительных предположений, осознанных или неосознанных. В частности, указание какой-то процедуры или алгоритма перехода от выборки с пропусками к совокупности скалярных оценок элементов фрагмента знаний — это пример неявного, а зачастую и неосознаваемого использования какого-то дополнительного предположения.

Мы будем, в основном, строить именно процедуры перехода и отмечать, где в них возникает «произвол» или «свобода выбора». Вопросы содержательной интерпретации сделанного выбора мы оставим в стороне, но подчеркнем, что они важны и пользуются заслуженным вниманием исследователей [1, 13–15].

Таблица 10

Выборки без пропусков, соответствующие (не противоречащие) выборке с пропусками

№	Выборки возможные							
	S_{90}	S_{91}	S_{92}	S_{93}	S_{94}	S_{95}	S_{96}	S_{97}
0	00	00	00	00	00	00	00	00
1	00	01	00	01	00	01	00	01
2	01	01	01	01	01	01	01	01
3	01	01	01	01	01	01	01	01
4	01	01	01	01	01	01	01	01
5	01	01	01	01	10	10	11	11
6	10	10	11	11	11	11	11	11
7	11	11	11	11	11	11	11	11
8	11	11	11	11	11	11	11	11
9	11	11	11	11	11	11	11	11
Пропущенные означивания	000	001	010	011	100	101	110	111
Означивание	Частота встречаемости означиваний							
00	2	1	2	1	2	1	2	1
01	4	5	4	5	3	4	3	4
10	1	1	0	0	1	1	0	0
11	3	3	4	4	4	4	5	5

Вернемся к табл. 6. Поскольку, вообще говоря, порядок следования означиваний в выборке нам не важен (*это* — одно из предположений), произведем их переупорядочение, чтобы было удобнее сравнивать между собой возможные выборки. Затем посчитаем частоту встречаемости означиваний в каждой их них.

Поделив частоту встречаемости каждого означивания на 10 (число элементов в выборке), мы получим для каждой выборки S_{9r} $(0 \le r \le 7)$ обученный вектор вероятностей квантов $\mathbf{Q}_{9r}$, а за счет преобразования $\mathbf{P}_{9r} = \mathbf{J}\mathbf{Q}_{9r}$ — обученный вектор вероятностей конъюнктов. Мы можем прибегнуть и к другим рассмотренным выше способам обучения этих векторов; главное — обученные векторы будут содержать скалярные оценки вероятностей. Мало того — они будут соответствовать непротиворечивым фрагментам знаний.

На данном этапе снова возникает возможность выбора. От совокупности векторов $\{\mathbf{Q}_{9r}\}_{r=0}^{r=7}$ требуется перейти к вектору $\mathbf{Q}_9$, который бы являлся результатом обучения по исходной выборке, содержащей означивания с пропущенными элементами.

Заметим, что у нас нет никаких априорных причин оказать предпочтение одним возможным выборкам вида S_{9r} перед другими. Если это так, то каждой из них можно сопоставить одинаковый неотрицательный вес так, чтобы сумма всех весов была равна 1. В рассматриваемом конкретном случае вес $w_r = \frac{1}{8}$ ($0 \le r \le 7$). Воспользуемся операцией линейной комбинации

$$\mathbf{Q}_9 = \sum_{r=0}^{7} w_r Q_{9r};$$

за счет равенства весов в данном случае получится

$$\mathbf{Q}_9 = \frac{1}{8} \sum_{r=0}^{7} Q_{9r},$$

более того, проведя необходимые подсчеты, мы получим:

$$\mathbf{Q}_9 = \begin{pmatrix} 0.15 \\ 0.40 \\ 0.05 \\ 0.40 \end{pmatrix}, \ \mathbf{P}_9 = \begin{pmatrix} 1 \\ 0.80 \\ 0.45 \\ 0.40 \end{pmatrix}.$$

Таким образом, перебирая все возможные выборки (не противоречащие исходной выборке с пропусками), предполагая, что они равноценны (равновероятны), мы смогли обучить векторы $\mathbf{Q}_9$ и $\mathbf{P}_9$ так, что у них оказались скалярные оценки. Более того, соответствующие им фрагменты знаний оказались непротиворечивыми.

У предложенной процедуры есть два очевидных недостатка. Во-первых, если пропусков много, то такая процедура, исполненная, как описано, будет вычислительно дорогостоящей: пространство возможных выборок растет экспоненциально от числа пропущенных элементов в означиваниях. Во-вторых, мы не можем быть уверены, что возможные выборки равноценны с точки зрения предметной области: могут существовать какие-то факторы, которые потребовали бы выбрать иное распределение весов, а не равномерное, которым мы воспользовались.

Второй недостаток можно компенсировать либо более глубоким анализом предметной области, чтобы оправдать выбор того или иного распределения весов, либо параллельно строить фрагменты знаний с интервальными оценками, по которым мы можем судить о степени произвола, допущенного при выборе одного распределения из многих возможных.

Приведем два примера сведений, которые повлияли бы на распределение весов в рассмотренной выше ситуации. Во-первых, совокупное число положительных означиваний на местах пропущенных элементов по каким-то причинам может быть ограничено 1 или 2. Тогда возможные выборки, не отвечающие этому ограничению, должны получить вес, равный нулю. Во-вторых, может быть известно вероятностное распределение возможных наборов означиваний пропущенных элементов. Например, в простейшем случае вероятность набора означиваний может определяться по закону Бернулли. Эти сведения также повлияли бы на веса возможных выборок. Заметим, что приведенные примеры отнюдь не исчерпывают все виды сведений, которые могут повлиять на выбор весов.

Недостаток, упомянутый выше первым, можно компенсировать за счет более тонкой организации процедуры перебора: чтобы получить окончательные оценки, от нас не требуется в явном виде строить весь набор возможных выборок. В некоторых случаях за счет этой оговорки можно редуцировать объем требующихся вычислений. В частности, мы можем воспользоваться методом Монте-Карло и сгенерировать не все 2^k выборок, а лишь какую-то их часть. Скорее всего, генерацию выборок потребуется продолжать до тех пор, пока вероятности в обученных векторах не стабилизируются с определенным уровнем точности.

9. Восстановление пропущенных значений

Кроме работы с возможными выборками, вполне оправданным было бы рассмотреть и альтернативный путь — каким-то образом восстановить пропущенные значения или разумным образом подменить означивания с пропусками набором полноценных означиваний, каждому из которых приписан вес. Сумма этих весов должна быть равна весу исходного означивания с пропуском; как мы помним, в выборке все означивания равноценны, поэтому вес каждого элемента можно посчитать равным 1 (или $\frac{1}{N}$, где N — число элементов в выборке, если соблюсти условие нормировки).

Разберем сначала конкретный пример. Пусть на основе уже проанализированных означиваний из выборки, которые были без пропусков, обучен вектор вероятностей квантов $\mathbf{Q} = \begin{pmatrix} p_{00} \\ p_{01} \\ p_{10} \\ p_{11} \end{pmatrix} = \begin{pmatrix} 0.15 \\ 0.40 \\ 0.05 \\ 0.40 \end{pmatrix}$. На вход поступает очередное означивание из выборки, которое содержит пропуск и имеет вид 1*. Ему не противоречат два означивания: 10 и 11.

Для замены (а значит, для восстановления пропущенного значения) поступившего 1* можно использовать означивание 11, как имеющее наибольшую вероятность. Такой подход приведет к контрастированию вероятностей элементов по принципу: растет вероятность того означивания, у которого она и так больше.

Но можно поступить иначе. Заменить 1* *набором* полных означиваний: внести в выборку означивание 10 с весом $\frac{0.05}{0.05+0.40}$ и означивание 11 с весом $\frac{0.40}{0.05+0.40}$. Этот подход будет поддерживать сохранение соотношений вероятностей. Следует заметить, однако, что на получаемый результат может повлиять порядок поступления элементов обучающей выборки.

В общем случае можно поступать одним из описанных выше способов. Либо выбирать из возможных полных означиваний, не противоречащих поступившему, то, которое наиболее вероятно, либо поступившее неполное означивание заменять набором полных означиваний, ему не противоречащих, с весами, полученными с помощью перенормировки из текущих обученных вероятностей этих элементов. В обоих случаях возникает проблема, когда алгоритм только начинает работать, а на вход уже поступает неполное означивание: час-

Труды СПИИРАН. 2008. Вып. 7. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2008. Issue 7. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

тотный подход еще не позволил накопить сведения о вероятностях. Чтобы преодолеть возникающую трудность, удобно задавать априорное распределение вероятностей на квантах. И это — еще один выбор, который влияет на результат обучения и который приходится делать.

10. Направления дальнейших исследований

В работе были рассмотрены подходы к локальному автоматическому обучению алгебраических байесовских сетей, то есть к формированию означиваний параметров фрагмента знаний (иначе — формированию численных оценок истинности конъюнктов или квантов) по выборке означиваний совокупности случайных бинарных элементов.

В случае исходных данных без пропусков на основе частотного подхода удается рассчитать скалярные оценки вероятностей. Расчет оценок может опираться либо на частотный подход, либо на подход, учитывающий не только частоту встречаемости в выборке, но и заданное априорное распределение. Формулы для обучения можно модифицировать также таким образом, чтобы оно стало адаптивным.

В случае исходной выборки с пропусками можно получить интервальные оценки вероятностей, не используя дополнительных предположений. Однако можно предложить процедуры расчетов, которые обеспечат нам получение скалярных оценок и в этом случае. Такие процедуры требуют расширенного исследования, поскольку опираются на дополнительные, зачастую неявные предположения, что требует впоследствии проверки применимости указанных процедур к конкретной ситуации.

Изложив основную идею локального автоматического обучения алгебраических байесовских сетей, отметим, что с точки зрения теории АБС остается неосвещенным ряд вопросов:

1) систематизация способов восстановления пропущенных значений и способов обработки выборки с такими значениями;
2) семантика этих способов (экспликация предположений, лежащих в их основе);
3) исследование свойств этих способов и результатов их применения;
4) возможные случаи дополнительных ограничений[2] на пространство возможных выборок, не противоречащих исходной выборке с пропусками;
5) разработка алгоритмов локального обучения; их анализ с точки зрения эффективности;
6) разработка приближенных алгоритмов, поскольку заранее можно ожидать, что «прямолинейные» «точные» алгоритмы окажутся вычислительно неэффективными;
7) принципы и алгоритмы выделения фрагментов знаний по доступной выборке;
8) согласование несовпадающих фрагментов знаний, которые имеют общие элементы, но обучались по разным выборкам;
9) разработка глобальных алгоритмов обучения, позволяющих построить и параметризовать алгебраическую байесовскую сеть, а не только отдельный ее фрагмент.

[2] В частности, могут быть доступны сведения, что может быть положительно означен только один литерал из каких-то зафиксированных двух, либо что из положительного означивания одного следует положительное означивание другого.

Наконец, заметим, что часть вопросов может быть уже решена в других разделах машинного обучения (например, для байесовских сетей доверия [14, 16]); такие решения требуется адаптировать для теории АБС, а часть вопросов ищет своих первых ответов.

*Часть результатов получена в рамках научно-исследовательских работ, осуществляющихся при финансовой поддержке РФФИ, проект **№ 09-01-00861-а.***

Литература

1. *Загоруйко Н. Г.* Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999. 270 с.
2. *Косовский Н. К., Тишков А. В.* Логики конечнозначных предикатов на основе неравенств. СПб.: Изд-во СпбГУ, 2000. 286 с.
3. *Тулупьев А. Л.* Алгебраические байесовские сети: теоретические основы и непротиворечивость. СПб.: СПИИРАН, 1995. 76 с.
4. *Тулупьев А. Л.* Алгебраические байесовские сети: логико-вероятностный подход к моделированию баз знаний с неопределенностью. СПб.: СПИИРАН; ООО Изд-во «Анатолия», 2000. 282 с.
5. *Тулупьев А. Л.* Алгебраические байесовские сети: локальный логико-вероятностный вывод. СПб.: СпбГУ; Анатолия. 2007. 80 с. (Элементы мягких вычислений.)
6. *Тулупьев А. Л.* Алгебраические байесовские сети: глобальный логико-вероятностный вывод. СПб.: СпбГУ; Анатолия. 2007. 40 с. (Элементы мягких вычислений.)
7. *Тулупьев А. Л.* Байесовские сети: логико-вероятностный вывод в циклах. СПб: Изд-во СпбГУ, 2008. 140 с. (Элементы мягких вычислений.)
8. *Тулупьев А. Л., Сироткин А. В.* Байесовские и марковские сети: логико-вероятностный вывод в базах фрагментов знаний с неопределенностью // Научн. Конф. Информационные технологии и системы, Геленджик, сентябрь 29 – октябрь 03, 2008 г.: Труды конференции. М.: ИППИ РАН, 2008. С. 440–456.
9. *Тулупьев А. Л., Столяров Д. М., Ментюков М. В.* Представление локальной и глобальной структуры алгебраической байесовской сети в Java-приложениях // Труды СПИИРАН. 2007. Вып. 5. СПб.: Наука, 2007. С. 71–99.
10. *Тулупьев А. Л., Николенко С. И., Сироткин А. В*. Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.
11. *Сироткин А. В., Тулупьев А. Л.* Матрично-векторные уравнения локального логико-вероятностного вывода в алгебраических байесовских сетях // Труды СПИИРАН. 2008. Вып. 6. СПб.: Наука, 2008. С. 131–149.
12. *Хованов Н. В*. Анализ и синтез показателей при информационном дефиците. СПб.: Изд-во С.-Петербургского ун-та, 1996. 196 с.
13. *Allison P. D.* Missing Data. London: Sage Publications, 2001. 93 p.
14. *Korb K. B., Nicholson A. E.* Bayesian Artificial Intelligence. NY.: Chapman and Hall/CRC, 2004. 364 p.
15. *Little R. J. A., Rubin D.* Statistical Analysis with Missing Data. 2nd ed. N.Y.: Wiley, 2002. 408 c.
16. *Perl J.* Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. NY etc.: Morgan Kaufmann Publ., 1994. 552 p.